Chapter Five:

Dynamical cognitive science and emotion

> *What is passionate in us rises and falls, leaps or creeps, and slowly paces. Now it becomes urgent, now hesitant, now stirred more feebly, now more strongly.*
>
> Johann Gottfried Herder

In the past decade, there have been a number of calls for cognitive scientists to adopt a more 'dynamical' approach to the study of the mind. In this chapter, I ask what such calls actually mean. In the second section, I examine how the dynamical approach can help to refine the interruption theory of emotion.

5.1. Dynamical cognitive science

The machines designed by cognitive scientists in the period 1950-1980 were almost all discrete-state machines. In such machines, the transitions between one state and another are like sudden jumps or clicks. There are no intermediate positions between one state and another.

Many of the first generation of cognitive scientists assumed, not only that this was the best way to design *artificial* minds, but that discrete-state machines would also be the best way to model *natural* minds. This assumption, however, is logically independent of both CTM and the design-based approach. The assumption of discreteness should not, therefore, be seen as an essential part of cognitive science *per se*, but simply as a characteristic feature of *classical* cognitive science. Its widespread acceptance among the first generation of cognitive scientists was a historical accident, not a logical necessity. This only became clear, however, when various sections of the cognitive science community began to propose other ways of building artificial minds.

The first cognitive scientists to do this were the connectionists. Rather than using a few heterogeneous components, each of which makes a specific isolable contribution to overall performance, connectionist models (often called nerual networks) employ a large number of homogeneous components. Whereas discrete-state machines usually process information serially, connectionist models are massively parallel (hence the term PDP, or 'Parallel Distributed Processing', that is often applied to such models).

The connectionists, however, were not always explicit about their commitments to continuous mathematics. Very often, they simply got on with the business of designing neural networks, and left the theoretical issues to philosophers. It was not until the emergence, in the 1990s, of proposals for a distinctively dynamical approach to cognitive science that the issue of continuity was really foregrounded.

Dynamical systems are not all continuous. There are discrete dynamical systems as well. But most proponents of the dynamical approach to cognitive science have tended to concentrate on continuous systems. It is therefore plausible to argue that the main contribution of the dynamical approach has been the challenge that it poses to the classical assumption of discreteness.

The dynamical hypothesis of cognition.

Dynamical cognitive science has been defined by its adherence to the ‘dynamical hypothesis of cognition’, which simply states that ‘cognitive agents are dynamical systems’ (van Gelder, 1998: 615). The term ‘dynamical system’ may be understood in various ways, but proponents of dynamical cognition usually construe it in very broad terms, as designating any system in which a variable, x, changes continuously over time. The ‘system’ in question may be (Smith, 1998: 4-5):

(i) *a real-world system*, such as the planets in motion or – in dynamical cognition – a cognitive agent, or

(ii) *a mathematical system*, which, in dynamical systems theory, may be (a) *a set of dynamical equations,* which perhaps aims to capture the behaviour of some real-world system, or (b) *an abstract mathematical structure*, such as a set of trajectories in a phase space which is characterised by a set of dynamical equations.

This multiple use of the term ‘system’ by the proponents of dynamical cognition is potentially harmful, as it might lead to ontological questions about the nature of things being confused with epistemological and methodological questions about the mathematical tools we use to describe and analyse things. To prevent such confusion, Tim van Gelder argues that we should distinguish two components of the dynamical hypothesis of cognition. On the one hand, ‘the *nature* hypothesis is a claim about the nature of cognitive agents themselves: it specifies what they *are* (i.e. dynamical systems)’ (van Gelder, 1998: 619, emphasis in original). The *knowledge* hypothesis, on the other hand, is a claim about cognitive science: namely, that we can and should understand cognition by using the resources of dynamical systems theory such as dynamical equations and geometrical modelling. Before going on to give some examples of research in dynamical cognition, I will briefly discuss the nature hypothesis and the knowledge hypothesis in turn.

The nature hypothesis

The nature hypothesis states that cognitive agents are dynamical systems (where ‘system’ is used in sense (i) described above). This purports to be an empirical claim about a set of real-world systems. However, things are not that simple. In particular, two points need to be made clear.

Firstly, it is rather misleading to claim that the statement ‘cognitive agents are dynamical systems’ amounts to a hypothesis about cognition, since this seems to imply that it makes some specific claim about cognitive agents. It sounds as if the property of being a dynamical system is being used to pick out cognitive agents from other kinds of physical entity. Yet this is not the case. *Any* physical system, from a single neuron to a galaxy, may be described as a dynamical system.[1] The concept of a dynamical system is no more specific than the rough idea of a computer as a system that transforms input into output. In fact, the two ideas are equivalent. To pick out those dynamical systems that are cognitive agents we need some further constraint, just as we did when we were trying to specify the notion of computation involved in CTM. We could, in fact, use the very same constraint that we used then – the notion of representation. We could, in other words, define cognitive agents as dynamical systems that represent other other dynamical systems.

On this account, dynamical cognitive science is thus no less ‘computational’ than other forms of cognitive science. Computers are just dynamical systems that represent other dynamical systems. When proponents of the dynamical approach, then claim that the view of the mind as a dynamical system ‘is an entirely different image from the brain as a computer with stored contents or subroutines to be called up by a program’ (Kelso, 1995: 1), they must clearly have in mind some much more specific notion of computation than the one I proposed in chapter one. They probably have in mind the classical view that defines computation as a process that terminates after a finite number of basic operations specified by an algorithm (in technical terms, an ‘effective’ process). This indeed, is Church’s thesis, which stipulates that a process could only be *called* 'computable' if it is effective in this sense. Church's thesis has acquired such standing in computational theory that it sometimes seems to be regarded as an empirical discovery, instead of the stipulative definition that it really is. Sterile disputes about whether or not the dynamical approach counts as ‘computational’ or not could be avoided

---

[1] To be more precise, any physical system can be construed as instantiating infinitely many dynamical systems. I come to this point shortly.

if it were kept in mind that the term can have two meanings. According to Church's thesis, computers are just finite-state machines.[2] According to the broader definition I proposed in the first chapter, finite-state machines are just one kind of computer; you can have continuous computers too.

The second point that I want to make about the nature hypothesis is this: to say that a real-world system, such as a convecting fluid, is a dynamical system is actually to make a claim about instantiation, not about simple identity. A convecting fluid is not itself a set of variables, but rather a material entity whose behaviour can be *described* by a set of variables. Likewise, cognitive agents are not sets of variables, but they may behave in ways that are describable in such terms. Thus the nature hypothesis should be understood as claiming that cognitive agents *instantiate* dynamical systems (van Gelder, 1998: 619).

Furthermore, since any real-world system instantiates *numerous* dynamical systems, the nature hypothesis should not be construed as claiming that each cognitive agent is some *particular* dynamical system. Rather, it should be construed as claiming that each cognitive agent 'is' as many dynamical systems as are needed to describe all the different kinds of cognitive performance exhibited by the agent (van Gelder, 1998: 619).

The fact that claims about the dynamical nature of cognition are really just claims about the various dynamical systems that cognitive agents instantiate threatens to undermine van Gelder's distinction between the nature hypothesis and the knowledge hypothesis. Claims about instantiation are reducible to claims about the theoretical tools that are most appropriate to use in studying a given object. The nature hypothesis thus reduces to the knowledge hypothesis.

The knowledge hypothesis

The knowledge hypothesis of dynamical cognition is much easier to state clearly than the nature hypothesis. It is simply the claim that cognitive agents are better understood by appealing to the resources of dynamical systems theory. Dynamical systems theory is a branch of pure mathematics concerned with the properties of dynamical systems (where 'system' is used in sense (ii) described above). Typically, dynamical systems theory uses a set of linked differential equations to specify the evolution in time of some variable, x. In such systems, time is continuous. But there are other dynamical systems, specified in terms of difference equations, in which time is discrete. So continuous dynamical systems are really just a subset of the class of all dynamical systems. However, for

---

[2] A finite state machine is a discrete state machine that consists of a *finite number* of discrete states, together with their state transition rules.

the reasons given above, this chapter deals only with continuous dynamical systems. To re-cap: a theory of minds as *discrete* dynamical systems is already available in the classical approach. The novelty of the dynamical approach consists, therefore, in providing an alternative theory of minds as *continuous* dynamical systems.

Still it remains to be seen how much of an ‘alternative’ this theory really amounts to. Some of the proponents of the dynamical approach write as if cognitive scientists are faced with a stark choice between it and the classical approach (e.g. van Gelder, 1998). What they mean, I suppose, is that cognitive scientists much choose between discrete and continuous models of cognition; the dynamical approach is not necessarily committed to rejecting the other assumptions of the classical approach (domain-generality and internalism). Yet the choice between discrete and continuous models is far from being a black-or-white one.

For a start, the distinction between discrete and continuous systems is a mathematical one, and it is not clear how to apply it to real-world systems. When a discrete model is criticised on the grounds that the real-world system is better represented by a continuous model, this can usually be reduced to one of the following two claims:

(i) The scale chosen by the discrete model is not sufficiently fine-grained: the dimension in question is usually (but not always) a *temporal* one.

(ii) The discrete model needlessly imputes too much internal structure to the real-world system.

However, since both of these criticisms can often be met by refining the discrete-state model, the real point at stake does not seem to be one continuity as such. I will now explain this point in more detail.

The grain problem.

Sometimes, I claim, the criticisms levelled against discrete models are reducible to the claim that the scale chosen by the discrete model is not sufficiently fine-grained. An example from the history of biology may serve to make this point clear.

In the first decades of the twentieth century, there was a rather silly feud between two ways of thinking about heredity. The Mendelians took a discrete approach; they argued that phenotypic characters were controlled by genes that were either present or absent (more precisely, only one particular *form* of a gene – one *allele* – could be present at any given chromosomal locus). The biometricians thought this was clearly at odds with the fact that some phenotypic characters can vary continuously; people are not, for example, just *tall* or *short*. Ronald Fisher showed that the disagreement could be solved if one assumed that such continuous phenotypic characters are polygenic – that is, controlled by more than one gene. If height is influenced by many genes of small effect, then it is clear how it can be approximately continuous.

The term ‘approximately’ is crucial. Whenever a discrete model differs from a continuous one, it can be made to approximate it more closely simply by giving it more discrete states. In other words, discrete models can always be made more realistic by choosing a more fine-grained scale of analysis. So when a discrete model is contrasted unfavourably with a continuous model, the charge often amounts to no more than the claim that the discrete model is not sufficiently fine-grained. And this criticism will only be valid if we can show that the scale chosen was not sufficiently fine-grained to meet the explanatory purpose for which the model was constructed. After all, the whole point of a model is not to be as fine-grained as the thing it is supposed to represent. It would be silly to criticise a map for not being as detailed as the terrain itself.

This point is borne out when one looks at the actual machines built by those who take a dynamical approach. These machines are usually just versions of connectionist networks. But connectionist networks are rarely built out of analogue components. They are usually simulated on discrete-state machines; the theoretically *continuous* activation level of each node in the network is, in practice, approximated by means of a fine-grained *discrete* series.

If connectionist networks can be implemented, to a good enough degree of approximation, on discrete-state machines, the reverse is also true. This, at least, seems to be what Fodor and Pylyshyn claim in their influential paper on connectionism and cognitive architecture (Fodor and Pylyshyn, 1988). They argue that a neural network architecture can implement the digital processes that they take to lie at the core of human cognition. If we combine these two points, we can imagine a machine that is digital at one level instantiating a system that, at a higher level of analysis, is continuous, and that this continuous system then instantiates a discrete-state machine at an even higher level. Thus the grain-problem does not just refer to the number of units on a particular scale; it also refers to the level of nature we are analysing (though perhaps this amounts to the same thing).

The dimension of discrete models which proponents of dynamical cognition most often focus on in their pleas for a finer grain of analysis is the temporal one. In other words, criticisms of discrete models often amount to no more than the charge that they have not been sufficiently sensitive to the details of timing. This may be a serious defect with some discrete models, but if so, it is not because they are *discrete*. It is just because they are not sufficiently aware that, for many cognitive processes, every millisecond counts. This problem can always be remedied by choosing smaller units for the temporal scale.

Parsimony

Whenever the criticisms levelled at discrete models cannot be plausibly construed as pleas for a finer-grain of analysis, they are usually reducible to the claim that the discrete models needlessly impute too much internal structure to the real-world system. To understand the connection between the idea of continuity and the idea of complexity, we need to make a brief digression into chaos theory.

Chaos theory is simply another name for the branch of dynamical systems theory that studies the properties of dynamical systems governed by nonlinear equations (Stewart, 1989: viii). An equation is linear if the sum of two solutions is itself a solution. The solution for a two-stone disturbance of a liquid surface, for example, is just the sum of the solutions for two one-stone disturbances, centred at appropriate points (Stewart, 1989: 72). Most classical equations, including those of classical dynamics, are linear. This is not true of the equations in chaos theory.

In chaotic systems, trajectories from nearby initial conditions can lead to outcomes that are not correlated. In other words, these systems exhibit sensitive dependence on initial conditions, a phenomenon that is sometimes known as ‘the Butterfly Effect’.[3] Some popularisations convey the impression that this is the most original ‘take-home message’ of chaos theory (Stewart, 1989), but this is misleading to say the least. It is no news that small changes can have huge effects; Darwin once remarked that his voyage on the *Beagle*, which determined his whole career, had at one point depended on such trifles as the shape of his nose.[4] The simple

---

[3] The name comes from a story which is often used in the literature to illustrate the phenomenon of sensitive dependence on initial conditions. Here is Peter Smith’s version of the story: ‘A small blue butterfly, let’s suppose, sits on a cherry tree in a remote province of China. As is the way of butterflies, while it sits it occasionally opens and closes its wings. It could have opened its wings twice just now; but in fact it moved them only once. And – because the weather system exhibits sensitive dependence – the miniscule difference in the resulting eddies of air around the butterfly eventually makes the difference between whether, two months later, a hurricane sweeps across southern England or harmlessly dies out over the Atlantic. Or so the story goes.’ (Smith, 1998: 1)

[4] Robert Fitzroy, the captain of the *Beagle*, was a firm believer in physiognomy, according to which a person’s character could be discovered by attending to the shape of their facial features. Darwin’s nose seemed to Fitzroy to betray a lack of resolution, and thus almost cost Darwin his place on the ship (Bateson and Martin, 1999: 123-24).

idea that small changes can cause large, unpredictable effects has been around for much longer than chaotic dynamics (Smith, 1998: 1).

Chaos theory gives to a new twist to this old idea by showing that the very complex patterns formed by the trajectories in a chaotic phase-space can be produced by relatively simple equations.[5] A very complex series, which may even appear to be completely random to the untrained eye, can be produced by a deterministic formula.

The take-home lesson here is that when we observe some very complex behaviour, we should not assume that the system producing it necessarily has a complex internal structure. Now, it is probably fair to say that many of the first cognitive scientists tended to make exactly this assumption. Guided by the strategy of functional decomposition (see section 1.1), they tended to assume that all complex behaviour could only be generated by a set of heterogeneous components, like those in, say, a television or an electronic computing machine *circa* 1960. One of the surprising things about connectionist models was that they put this assumption in doubt. Some kinds of task, such as pattern recognition, could be achieved by very simple networks consisting of a few dozen homogenous components (simple nodes).

Hype

If I am right, and most (perhaps *all*) criticisms levelled at discrete models can be reduced either to a plea for a finer grain of analysis or a call for greater parsimony, then the problem with such models is not their discreteness. Such criticisms could be met without making the model continuous. In neither case is the point at stake one of continuity as such.

Nor is it strictly necessary to appeal to the resources of dynamical systems theory to make these points. The point about choosing the appropriate grain of analysis can be made quite well without them. And, as the example of connectionism shows, it is not necessary to invoke the arcane terminology of chaos theory to make the point that complex behaviour can sometimes be generated by relatively simple systems. The connectionists did not often describe their models in terms of chaotic attractors, but they succeeded in challenging the assumption of internal complexity that seemed to underlie many classical models.

---

[5] There seems to be some sleight of hand here. Unless we can specify some uniform measure of 'complexity' that applies equally to mathematical equations and to the patterns formed in a multi-dimensional phase-space, then it makes no sense to marvel at how wonderfully 'complex' instances of the latter can be produced by such 'simple' instances of the former.

One is left with the impression that the real value of dynamical systems theory for cognitive science is rhetorical. The arcane mathematical terminology is being used not for any intrinsic value, but merely to make people think that there *is* something important being said here. Now, sometimes there is something important. Sometimes the discrete models do need to be finer-grained and more parsimonious. But these points would, I think, be made more persuasively if they were couched in more simple terms.

Kelso and other dynamicists are well aware of these suspicions, and try to distance themselves from the hype and the vague analogies that have characterised some of the recent writings on chaos by insisting on precise links between theory and data:

> I am not going to comment on the rhetoric surrounding the buzzword chaos and how it provides a more holistic view of human life, except to say, chaos of what? What are the relevant variables that are supposed to exhibit chaotic dynamics? What are the control parameters? And how do we find them in complex living systems where many variables can be measured, but not all are relevant?...What are the attractors? What does the bifurcation diagram look like? Are these concepts and mathematical tools even relevant? How does one establish them, even in a single case?...There has to be some connection between mathematical formulae and the phenomena we are trying to understand....Establishing a connection between theory and experiment is one of the canons of science that the ‘chaos, chaos everywhere’ crowd seems to ignore.
>
> (Kelso, 1995: 43-44)

Kelso rightly draws importance to the importance of empirical research, and his book includes some of the best examples of such work. One such example is the Haken-Kelso-Bunz (HKB) model of limb coordination. Kelso and colleagues studied coordination within and between limbs, and found that a whole range of different coordination patterns could be modelled with the same dynamical equation. The only variable required for this equation was the relative phase of the limbs in question. Relative phase refers to the relation between two oscillating components. Imagine that you are tapping the table in front of you with the forefinger of each hand. If the fingers hit the table simultaneously each beat, the relative phase is said to be ‘inphase’, while if the fingers hit the table alternately, the relative phase is ‘antiphase’.

Kelso also found that a whole variety of limb coordination patterns, from the case of finger-wagging just described, to trotting and galloping in quadrapeds, could be modelled by a dynamical equation based on the derivative of the relative phase. This simple model predicted a wide of range of observed phenomena, including the small range of stable coordination patterns and the nonequilibrium phase transisitions that occurred when the system moved from one stable pattern to another (Kelso, 1995: 46-57, 74-87).

The limited applicability of the dynamical approach

The empirical work described by Kelso, such as the HKB model, goes some way towards pre-empting the charge of 'hype' But finger wagging and limb coordination are not exactly paradigms of 'mental' processes. Ture, movement is vital to all natural cognitive agents, but classical cognitive science did not completely disregard motor control. On the other hand, classical cognitive science also provided models for more paradigmatic mental processes like reasoning and problem-solving, but dynamical models for such things are practically non-existent. Even if we grant that there are natural advantages to modelling motor-control in continuous terms, it seems hard to believe that we could say the same about other mental processes.

Evolutionary cognitive science can suggest a very good reason for this. Natural minds evolved to guide adaptive behaviour; and it would almost never be useful to have cognitive or emotional systems do large numbers of iterations through dynamical states into order to achieve such behaviour.[6] Rhythmic locomotion seems to be the only case where coupled oscillators have a distinct advantage over discrete-state machines. No wonder, then, that dynamical models of 'cognition' tend to concentrate on things like limb-coordination.

Even if we include connectionist networks in the class of continuous systems, the dynamical approach does not go much further. The only area in which connectionist networks seem to have a distinct advantage over classical discrete-state machines seems to be in the area of pattern recognition. Insofar as continuous models can be taken as a separate class from discrete models, then, their value may be restricted to 'low level' processes such as perception and motor-control.

Is the dynamical approach design-based?

Proponents of the dynamical approach point out that connectionism is not the same thing as dynamics (van Gelder, 1998: 661). But this only weakens their claim to make a distinct contribution to cognitive science. The distinctive thing about cognitive science, I argued in chapter one, is its design-based methodology. If the dynamical approach cannot claim connectionist networks as paradigms of the continuous models they

---

[6] I owe this point to Geoffrey Miller (personal communication).

prefer, then it the dynamical approach is reduced to mere theory without any concrete proposals to show in the way of working machines. If the dynamical approach is to prove itself as a species of cognitive science, it must provide clear evidence of how it leads to new insights in artificial intelligence. So far, it has failed to do so. The various 'dynamical machines' all turn out to be old-fashioned connectionist networks, with perhaps the twist that several such networks are linked up in novel ways (see, for example, the robot described in Tani, 1999: 157).

Besides, even if we grant, for the sake of argument, that connectionist networks *are* distinctively dynamical machines, it is still not clear whether this will permit the dynamical approach to call itself design-based. The point of the design-based methodology is not simply that we build working machines, but that we understand *how* the machines work. With traditional digital machines, this is no problem. One can apply the strategy of functional decomposition quite easily to such machines. But connectionist networks are no so amenable to this explanatory strategy. There is no obvious way of assigning different functional systems such as 'memory' or 'executive' to a particular bit of the network. Rather, these functional systems tend to be 'distributed' across the whole network in a way that defies clear explanation. Connectionist networks thus seem to be much less 'transparent' in their workings than digital machines. This may be why people turn to mathematical equations to understand them; there is simply no other way. But describing a machine in terms of the mathematical function it computes is rather different to breaking it down into distinct subsystems. To the extent that the design-based approach of cognitive science requires functional decomposition, connectionism does not count as part of cognitive science.

5.2. Dynamical approaches to emotion

In section 3.2, I proposed, on the basis of various evolutionary hypotheses, a definition of emotions as interruption mechanisms. In section 4.2, I argued that we could take this theory further by drawing on the insights of situated cognitive science. Can we take it further still by drawing on the insights of the dynamical approach?

These insights, I have argued, reduce to two points. To recap, if the dynamical approach offers anything distinctive, it is a reminder that we should be careful not get the grain of our analysis right, and not to assume that complex internal structure is necessary to generate complex behaviour. Let us see how we can use these insights to refine the interruption theory of emotion.

Are there enough states in the model?

The classical models of emotion described in chapter two, such as the OCC model, represent emotions in discrete terms. In response to a given input, a description of an emotion is either generated or it is not. There are no intermediate states between the full presence of an emotion and its complete absence.

In humans and other animals, however, emotions are not such black-and-white affairs. They have quantitative characteristics as well as qualitative ones. At any one moment, an emotion may be present in varying degrees of intensity, and this intensity may wax or wane with varying rapidity. These quantitative aspects of emotion can be modelled in a discrete-state machine by adding more states. Rather than just assigning one bit to emotion (is it on or off?), we can assign several, representing degrees of emotional intensity. Or we can use a more obviously analogue system, such as a connectionist network.

Juan Velasquez of MIT has developed a connectionist model of emotion called ‘Cathexis’ (Velasquez, 1996). The network consists of three layers of nodes. In the input layer, the nodes represent the four kinds of emotional elicitor listed by Caroll Izard in her theory of emotion: neural, sensorimotor, motivational and cognitive (Izard, 1993). In the middle layer, each node represents an emotion such as joy or distress. In the output layer, each node represents a behaviour, such as smiling. The nodes in the middle layer are connected to each other as well as to the nodes in the other layers, so, for example, joy can inhibit distress and activate hope. Each emotion-node has a continuously variable level of activation, which represents the intensity of that emotion at a given point in time. Unlike the classical models, then, in which only one emotion may be present at a given time, all emotions are constantly activated in Cathexis, though at varying levels of intensity. The intensity of each emotion changes at regular intervals in accordance with an equation whose terms include the inputs from other nodes modified by inhibitory and excitatory gains, and a function that controls the temporal decay of the emotion. The new intensity is thus a function of its decayed previous value, the effects of its elicitors, and the influences of other emotions.

Cathexis manifests the quantitative features of emotion that the classical models leave out. Emotions are not simply present or absent, but are continuously present in varying degrees of intensity. Furthermore, the intensity of each emotion changes at differing rates. The equation that specifies how the intensity changes from one moment to the next is nonlinear, so chaotic dynamics may be observed. For example, since the equation includes parameters that specify a minimum activation threshold and a maximum saturation value, the possible intensity values for each emotion may be graphed as a sigmoidal curve in which the middle region is highly sensitive to initial conditions. The steepness of the sigmoid can be altered by changing the values of these parameters, which allows for different temperaments to be modelled. A steep sigmoid, for example, would reflect an emotionally labile temperament, while a gentler slope would represent a more phlegmatic character. Finally, the multiple interactions among the nodes allows for feedback loops with various nonlinear properties such as time-dependence (i.e. the effect of an elicitor on the intensity of an emotion depends on the time when the elicitor comes into play).

The interruption theory could learn from this model. I have already mentioned the possibility that the top-down effects of cognition on emotion might be modelled by allowing the top layer in the hierarchy (the cognitive layer) to have some control over the activation threshold of the lower layers (the interruption mechanisms). This is already to introduce the idea of variable intensity into our model. Cathexis suggests that we might extend this idea to interactions between the lower layers themselves. If the activation threshold of the lower layers was not influenced just by the top layer, but also by other lower layers, then feedback effects among the various layers could emerge. Increasing tiredness could, for example, lower the activation threshold for anger; this would model the tendency for tired people to lose their temper more easily than others.

Whether or not the addition of these quantitative features should be seen as inherently non-classical is dubious, however. They could be built into the interruption model without using a connectionist architecture, simply by using discrete-state machines with many more states to approximate the continuously variable intensity of emotions. Even if we did build some of the layers along connectionist lines, we would not need to build all the layers along such lines; we could use a hybrid architecture. And even if the machine were entirely composed of connectionist networks, it would still have features that could be described in discrete terms. Cathexis, for example, though touted as a thoroughly dynamical model of emotion, has some discrete aspects. Thus, while the intensity of each emotion appears to vary continuously, this is only because the discrete series of intensity values has many members. Furthermore, the fact that each emotion node only excites other nodes when its intensity surpasses a given threshold bestows a digital character on excitation.

Mood

Building variable activation thresholds into the interruption theory may provide us with a good way of understanding what moods are. The features that are usually used to distinguish between emotions and moods are all phenomenological rather than mechanistic. Moods are usually said to build up and die away more slowly than emotions, to last longer than emotions, and to constrain attention less forcefully than emotions . If this is all there is to the emotion-mood distinction, we need not regard moods as separate mechanisms from emotions but merely as a set of dimensions along which various features of the emotion mechanisms can vary. An irritable mood, for example, could be modelled by simultaneously altering various parameters of the anger mechanism: for example, we might lowering the activation threshold, decreasing onset time, and increasing offset time. Let us call this the 'parameter theory of mood'.

This way of understanding mood would actually correspond quite well with the explication of mood offered by Vincent Nowlis. Nowlis suggested that moods were second-order dispositions (Nowlis, 1963). In other words, while emotions are dispositions to act in particular ways, moods are dispositions to have certain emotions. The parameter theory of mood provides a concrete way of understanding this rather abstract definition. It gives a precise mechanical account of how the dispositions are realised in design terms. It also accords well with the common view of moods as 'emotional *states'*. That is, if emotional episodes are by definition relatively short-lived, this is because such episodes are to be identified, in our model, with the relatively short period during which a lower-layer takes control of behaviour. Whenever, the top layer is in control, the system is not 'in the grip of' an emotion. However, the system is always in some particular emotional state or other, in the sense that we can always give an answer to the question: 'which emotional mechanism has the lowest activation threshold at the moment?' The system is always in one mood or another, even when none of the interruption mechanisms are in control of behaviour. For example, if the anger mechanism has the lowest activation threshold, then we say that the system is in an 'irritable mood'.

Feedback

By allowing for feedback effects between the activation thresholds of the interruption mechanisms, we could also model the oscillations of mood that are commonly observed in normal people and which are more pronounced in those suffering from various mood disorders. There might even be a role for coupled oscillators here, which would increase the relevance of a dynamical approach to the mind by showing another role for oscillators in the mental economy other than rhythmic locomotion. Perhaps unipolar mood disorders could be modelled as point attractors in the phase space of emotional states, and bipolar disorders as limit cycles. The dynamics of mood disorders are not well understood, and such a model could prompt us to look for relevant data by suggesting possible control parameters.

Does the interruption model impute too much internal structure to emotional mechanisms?

In the previous section, I argued that many criticisms of discrete models can be reduced to the claim that the discrete model imputes too much internal structure to the real-world system. In cognitive science, this amounts to the charge that discrete models sometimes present a rather baroque view of the mind which flouts the principle of parsimony.

It is hard to tell whether or not the interruption theory could be accused of attributing too much internal structure to the interruption mechanisms, since I have not offered any detailed account of how these mechanisms are supposed to work. Pending such accounts, we must suspend our judgement on this question. However, when constructing design hypotheses for these mechanisms, we should bear in mind the general principle that complex behaviour can sometimes be generated by relatively simple mechanisms.

These rather bald statements do not go very far towards demonstrating the usefulness of adopting a dynamical approach to emotion. It remains to be seen, therefore, whether or not the dynamical approach can contribute as much to interruption theory as the other non-classical approaches outlined in this thesis.